NATIONAL TRANSPORTATION SAFETY BOARD
Office of Aviation Safety
Washington, D.C. 20594

December 22, 2003

Human Performance

SPEECH STUDY REPORT

A. ACCIDENT

Operator: American Airlines (flight 587)
Location: Belle Harbor, New York
Date: November 12, 2001
Time: 0916 Eastern Standard Time
Aircraft: Airbus A300-600, N14053
NTSB Number: DCA02MA001

B. HUMAN PERFORMANCE GROUP

Malcolm Brenner, Ph.D.
National Transportation Safety Board
Washington, D.C. 20594

Doug Brazy
National Transportation Safety Board
Washington, D.C. 20594

C. ACCIDENT SUMMARY

On November 12, 2001, about 0916 Eastern Standard Time, American Airlines flight 587, an Airbus A300-600, was destroyed when it crashed into a residential area of Belle Harbor, New York, shortly after takeoff from the John F. Kennedy International Airport (JFK), Jamaica, New York. Two pilots, 7 flight attendants, 251 passengers, and 5 persons on the ground were fatally injured. Visual meteorological conditions prevailed and an instrument flight rules flight plan had been filed for the flight destined for Santo Domingo, Dominican Republic. The scheduled passenger flight was conducted under Title 14 Code of Federal Regulations (CFR) Part 121.

D. DETAILS OF THE STUDY

Speech evidence has been examined in previous investigations for potential evidence on human performance issues (References 1, 2, 4, 6, 9-13, 16). Similar efforts were made in the present investigation using audio information from the hot microphone cockpit voice recorder (CVR) channels that captured speech through boom microphones attached to headsets worn by each pilot. Subjective evaluations and computer scorings of the audio information were made to examine human evidence in three areas: non-verbal sounds, physical straining, and psychological stress.

D.1 Non-verbal sounds

The CVR transcript[1] lists several instances of non-verbal human sounds captured over the course of the recording. As shown in Table 1, which summarizes all such sounds in the CVR transcript, there were no non-verbal sounds identified for the captain while there were varied sounds, including four instances of yawning sounds, identified for the first officer.

D.2 Physical straining

The CVR transcript notes that the first officer's statement "max power" (0915:54.2) was "spoken in a strained voice." Since straining sounds might reflect an unusually strong physical effort by the pilot, and since this statement occurred during a period of particular importance to the investigation, the Human Performance Group examined this statement and surrounding statements to judge whether straining sounds might provide a profile of the first officer's physical efforts.

[1] Cockpit Voice Recorder Transcript, American Airlines Flight 587, DCA02MA001, Belle Harbor, NY, November 12, 2001. Exhibit 12A, Docket SA-522.

Meeting in the NTSB Audio Laboratory with support from the CVR Group Chairman (Reitan), Human Performance Group members reviewed every recorded statement on the first officer's hot microphone channel during the takeoff and wake encounter portions of the CVR tape (0913:21 to 0915:57.5).[2] For comparison purposes, the members listened repeatedly to the first officer's strained speech in saying the word "power" in the statement "max power" (0915:54.2) compared to his routine speech in saying the word "power" in the statement "climb power" (0914:56.5). Based on this comparison, each Human Performance Group member individually rated the level of strain perceived in the first officer's speech during each statement of the CVR transcript.

Group members agreed that there was evidence of perceived strain in the statements "yea, I'm fine" (0915:55.3) and "let's go for power please" (0915:57.5) that was similar to that in the statement "max power" but to a slightly lesser degree and perhaps varying over the course of the statements. There was no consensus on the exact profile of perceived strain in these statements. Group members agreed that there was a suggestion of strain in the statement "you got throttles" (0914.03.8) and no perception of strain in the remaining CVR statements. There was consensus among group members that it would be helpful to develop an objective measure of strain in speech as a future investigative technique for examining pilot actions.

D.3 Psychological stress

Research has shown that fundamental frequency ("pitch")[3] and amplitude ("loudness")[4] often convey information about a speaker's psychological stress (References 3, 5, 7, 8, 15, 18). With regard to fundamental frequency, the following guidelines (12-13) have been used to evaluate the approximate degree of psychological stress experienced by a speaker and its effect on performance. These guidelines are advisory and considered with other speech factors in characterizing the speaker's level of stress:

- An increase in fundamental frequency of about 30 percent (compared with that individual's speech in a relaxed condition) would be characteristic of a stage 1 level of stress, which could result in the speaker's focused attention and improved performance.

---

[2] No review was made of the final portion of the tape because of a marked change in recording quality and background noise after presumed tail separation.

[3] Fundamental frequency is the rate at which the vocal cords of the larynx open and close during speech releasing puffs of air. A fundamental frequency of 150 Hz. indicates that the vocal cords open and close 150 times per second. Listeners normally perceive the fundamental frequency as the pitch of the speaker's voice.

[4] Amplitude is the physical energy of the sound as produced by the amount of air moved through the vocal cords and normally measured in volts (shown here as a relative measure). Listeners normally perceive amplitude as the loudness of the speaker's voice.

- An increase in fundamental frequency of about 50 to 150 percent would be characteristic of a stage 2 level of stress, which could result in the speaker's performance becoming hasty and abbreviated and therefore degraded; however, the speaker's performance would likely not display gross mistakes.
- An increase in fundamental frequency of about 100 to 200 percent would be characteristic of a stage 3 level of stress, or panic, which would likely result in the speaker's inability to think or function logically or productively.

With regard to amplitude, there is evidence that similar increases occur with stress (Reference 5) although evaluation standards have not been established here.[5]

For the present investigation, computer measures of fundamental frequency and amplitude were made for every CVR statement by each pilot to examine whether these variables would provide meaningful evidence concerning stress. As a validation test of whether these pilots responded to stress with speech changes, statements were grouped according to stage of flight. All hot microphone statements recorded from 0846:50 until 0858:32 were classified as "parked," all statements recorded from 0859:34 until 0912:40 were classified as "engine start/taxi," all recorded from 0913:21 (Hot 1: "you have the airplane") until 0915:28.5 were classified as "takeoff," and all recorded from 0915:44.4 until 0915:57.5 were classified as "wake encounters."

Extraction of the speech sample

The CVR provided an audio record of the last 30 minutes, 39 seconds of the flight as recorded at a nominal rate of 1 7/8 inches per second on a Fairchild Model A-100 analog tape recorder located in the tail of the airplane. All speech samples in the study were taken from the hot microphone CVR channels that captured speech through boom microphones attached to headsets worn by each pilot and positioned in front of the mouth.[6] Recording quality was excellent.

The NTSB Audio Laboratory prepared digital copies of the audio recordings sampled at 32 KHz. For analysis, each statement was analyzed through the Entropic Signal Processing System (Entropic Research Laboratory,

[5] Amplitude is a difficult measure to interpret in operation settings because, unlike fundamental frequency, it is directly affected by factors such as distance between the speaker's mouth and the microphone, automatic gain features in the recording device, and ambient background noise. The hot microphone recordings make amplitude measurement especially usable because the boom microphones maintain a fixed distance to the speaker's mouth and are not subject to automatic gain control. However, background noise remains a significant potential confound, and amplitude is considered a secondary measure in this report.

[6] The first officer's hot microphone channel recorded speech throughout the CVR tape, but the captain's hot microphone channel did not begin recording speech until about 7 minutes into the CVR tape (0852:44) when the captain presumably donned the boom microphone as the cockpit door was closed.

Inc.) for computer-generated measures of fundamental frequency and amplitude. The program estimated the probability of voicing[7] at regular points within the waveform, and, for voiced samples, selected fundamental frequency estimates from candidates proposed by solving for the roots of the linear predictor polynomial computed periodically for the waveform. Amplitude was the estimate of the energy of the fundamental frequency. Speech samples were played over audio headphones as analyzed and were portrayed on a Waves digital video display along with computed analysis parameters calculated by the computer allowing the operator to monitor the analysis procedure. The operator entered missing data values whenever there appeared to be sufficient background noise or interference from another speaker to render the measurements uncertain (missing data was noted for 2% of the fundamental frequency statements and 9% of amplitude statements).

Results

Average speech measures were examined during the different stages of flight to assess whether the accident pilots responded to presumed increases in stress with corresponding increases in the speech measures. If so, a more detailed inspection was made of individual speech scores in an attempt to better understand individual pilot reactions during the flight. It was assumed, following previous literature using measures such as heart rate and electrodermal activity (References 14, 17, 19), that pilots would be most relaxed when the airplane was parked at the gate, that their stress level would increase as they began taxi operations, and that stress would be greatest during takeoff (prior to structural damage).

Figure 1 shows the average fundamental frequency scores for each pilot according to stage of flight. The plotted values have been normalized, so the two pilots can be directly compared. The sample sizes of each stage, reflecting the number of statements made during this stage corrected for missing data, were as follows: parked, 21 statements for the captain and 38 for the first officer; engine start/taxi, 42 for the captain and 39 for the first officer; takeoff, 15 for the captain and 12 for the first officer; and wake encounters, 3 for the captain and 6 for the first officer.

As shown in Figure 1, both pilots appeared to respond with characteristic changes in speech fundamental frequency to the increasing demands of different stages of flight from pre-taxi, through taxi, to takeoff. However, the First Officer showed a large response to the wake encounters that was not paralleled by results for the captain. To explore this difference, Figures 2 and 3 respectively

[7] Voicing refers to sounds produced by a periodic motion of the vocal cords, contrasted with unvoiced sounds which are produced by blowing air through a narrow opening (such as between the tongue and the roof of the mouth). Fundamental frequency and amplitude measures applied only to voiced samples.

plot the average fundamental frequency values for each statement made by the captain and first officer during the takeoff/wake encounter periods. For comparison, these figures display the level lines of percentage increase above a "relaxed" speech baseline (based on the average fundamental frequency value of 160.5 Hz. for the captain and 152.6 Hz for the first officer displayed during conversation when the airplane was parked). The "+30% above baseline" line provides a guideline for a stage 1 level of stress, the "+50% line provides a guideline for the beginning of stage 2, and the "+100%" line provides a guideline for the beginning of stage 3.

As shown in Figure 2, the captain displayed a relatively uniform profile of stress that remained within a generally alerted level (stage 1). As shown in Figure 3, the first officer showed a progressively increasing fundamental frequency in his last three statements that reached a high stress level associated with degraded performance (stage 2) although still remaining below a level associated with panic (stage 3). The last three statements showed fundamental frequency values that were significantly higher than those of all earlier statements by the first officer. [8]

Figure 4 shows the average amplitude scores for each pilot according to stage of flight. The plotted values have been normalized as in Figure 1, and the samples sizes were as follows: parked, 21 statements for the captain and 38 for the first officer; engine state/taxi, 31 for the captain and 38 for the first officer; takeoff, 15 for the captain and 12 for the first officer; and wake encounters, 3 for the captain and 5 for the first officer.

As shown in Figure 4, both pilots appeared to respond with characteristic changes in speech amplitude to the increasing demands of different stages of flight (Figures 5 and 6). The First Officer showed a progressive increase in amplitude during his last three statements.[9]

Submitted By:

____

Malcolm Brenner, Ph.D.
National Resource Specialist--Human Performance

[8] These last three statements were significantly higher statistically than previous statements in the distribution of the first officer's scores; respective z scores =2.28, p<.05; z=3.65, p<..001; z=4.63, p<.001.

[9] These last three statements were statistically higher than previous statements; respective z scores =1.85, p<.10; z=1.97, p<.05; z=2.21, p<.05.

REFERENCES

(1) Australian Bureau of Air Safety Investigation (1988). Accident Investigation Report, Mid-air collision between Cessna 172-N VH-HIZ and Piper PA38-112 VH-MHQ, near Tweed Heads, New South Wales, 20 May 1988 (BASI Report 881/1042)

(2) Australian Transport Safety Bureau (1999). Investigation Report 199902928, Inflight pilot incapacitation 72 miles east of Edinburgh, Aerodrome, Raytheon Aircraft, Super King Air 200, 21 June 1999. ATSB, Department of Transport and Regional Services, Government of Australia.

(3) Brenner, M., Doherty, E.T., Shipp, T. (1994). Speech measures indicating workload demand. *Aviation, Space, and Environmental Medicine, 65*:21-6.

(4) Brenner, M., Cash, J.R. (1991). Speech analysis as an index of alcohol intoxication—the Exxon Valdez accident. *Aviation, Space, Environmental Medicine, 62*: 893-8.

(5) Brenner, M.; Mayer, D.; and Cash, J. (1996). Speech Analysis in Russia. *Methods and Metrics of Voice Communications*. Ed. B. G. Kanki and O. V. Prinzo. Department of Transportation, Federal Aviation Administration, and Office of Aviation Medicine. DOT/FAA/AM-96/10.

(6) Government of Japan (1987). *Japan Air Lines Co., Ltd., Boeing 747 SR-100. JA8119, Gunma Prefecture, Japan, August 12, 1985.* Aircraft Accident Investigation Report, Ministry of Transport, Government of Japan, June 19, 1987 (English translation prepared by the Japanese government).

(7) Johannes, B., Salnitski, V.J., Gunga, H.-C., Kirsch, K. (2000). Voice stress monitoring in space—possibilities and limits. *Aviation, Space, Environmental Medicine, 71*: A58-65.

(8) Mayer, D.L., Brenner, M., Cash, J.R. Development of a speech analysis protocol for accident investigation. In *Methods and Metrics of Voice Communications.* Ed. B.G. Kanki, O.V. Prinzo. Washington, D.C.: Department of Transportation, Federal Aviation Administration, Office of Aviation Medicine, DOT/FAA/AM-96/10.

(9) National Transportation Safety Board (1990). *Grounding of the U.S. Tankship Exxon Valdez on Bligh Reef, Prince William Sound near Valdez, Alaska, March 24, 1989.* Marine Accident Report NTSB/MAR-90/94. Washington, D.C.

(10) National Transportation Safety Board (1992). NTSB Accident FTW91FA144, involving a Piper Arrow airplane at Midland, Texas, on 8/9/91; Speech Analysis Report, May 8, 1992.

(11) National Transportation Safety Board (1996). NTSB Accident SEA95FA175, involving a Beech F-33A airplane at Swan Lake, Montana, on 8/7/95; Factual Addendum, July 3, 1996.

(12) National Transportation Safety Board. (1999). *Uncontrolled Descent and Collision with Terrain, USAir Flight 427, Boeing 737-300, N513AU, near Aliquippa, Pennsylvania, September 8, 1994.* Aircraft Accident Report, NTSB/AAR-99/01, Washington, D.C., pp. 135-145.

(13) National Transportation Safety Board. (2001). *EgyptAir Flight 990, Boeing 767-366ER, SU-GAP, 60 Miles South of Nantucket, Massachusetts, October 31, 1999.* Aircraft Accident Brief, NTSB/AAB-02/01, Washington, D.C., pp. 21-25.

(14) Roscoe, A. H. (1992). Assessing pilot workload. Why measure heart rate, HRV and respiration? Biological Psychology, 34:259-287.

(15) Ruiz, R., Legros, C., Guell, A. (1990). Voice analysis to predict the psychological or physical state of a speaker. *Aviation, Space, and Environmental Medicine, 61*:266-71.

(16) Saito, I., Fujiwara, N., Utsuki, N., Mizumoto, C., Arimori, T. (1980). Hypoxia-induced fatal aircraft accident revealed by voice analysis*. Aviation, Space, and Environmental Medicine, 51(4)*: 402-406.

(17) Speyer, J.J., Blomberg, R.D., Fouillot, J.P. (1990). Evaluation the impact of new technology cockpits: onwards from A300FF, A310, A320 to A330, A340. Paper presented at the International Conference, Human Machine Interaction and Artificial Intelligence in Aeronatuics and Space, Toulouse, France. Reprinted in the Airbus Industrie Publication, *Workload and Vigilance, Physiologic Ambulant Monitoring, Automation and Error Tolerance, Cockpit Resource Management.*

(18) Williams, C.E., Stevens, K.N. (1981). Vocal correlates of emotional states. In: Darby, J.K., Ed. Speech evaluation in psychiatry. New York, NY: Grune & Stratton.

(19) Wilson, G. F. (2002). An analysis of mental workload in pilots during flight using multiple psychophysiological measures. International Journal of Aviation Psychology, 12(1): 3-18.

Table 1. Summary of non-verbal human sounds identified on the cockpit voice recorder (CVR) transcript.

CAPTAIN:

None

FIRST OFFICER:

0846:08 Sound of yawn
0846:51 Sound of chuckle
0850:21 Sound of hiccup and cough
0851:03 Sound similar to yawn
0851:11 Sound of singing
0851:24 Sound of chuckle
0855:23 Sound of humming
0909:27 Sound of yawn
0915:28.5 Sound similar to yawn
0916:00.0 Sound similar to human grunt

FIGURE 1

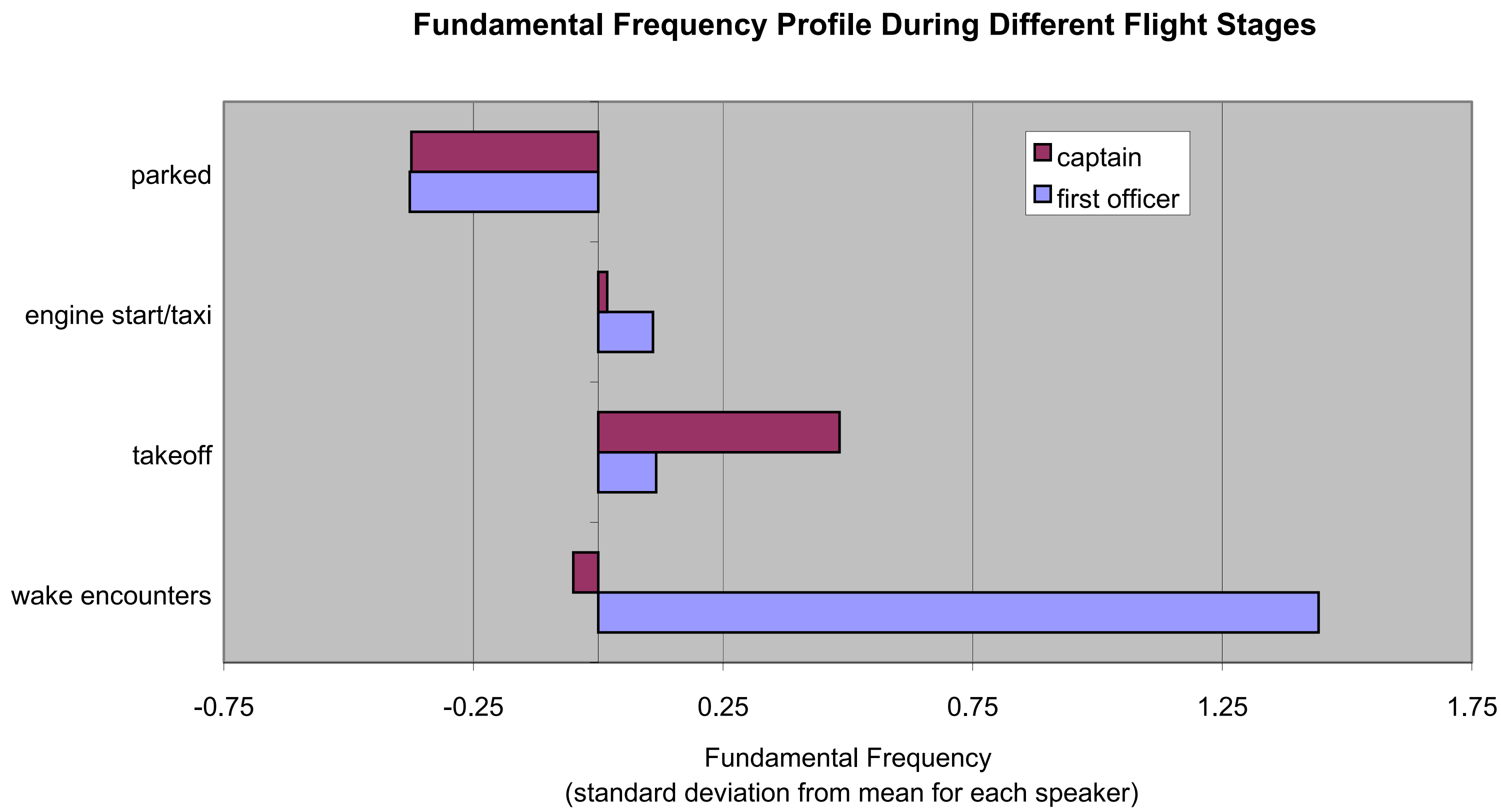

Fundamental Frequency Profile During Different Flight Stages
parked
engine start/taxi
takeoff
wake encounters
captain
first officer
-0.75
-0.25
0.25
0.75
1.25
1.75
Fundamental Frequency
(standard deviation from mean for each speaker)

FIGURE 2

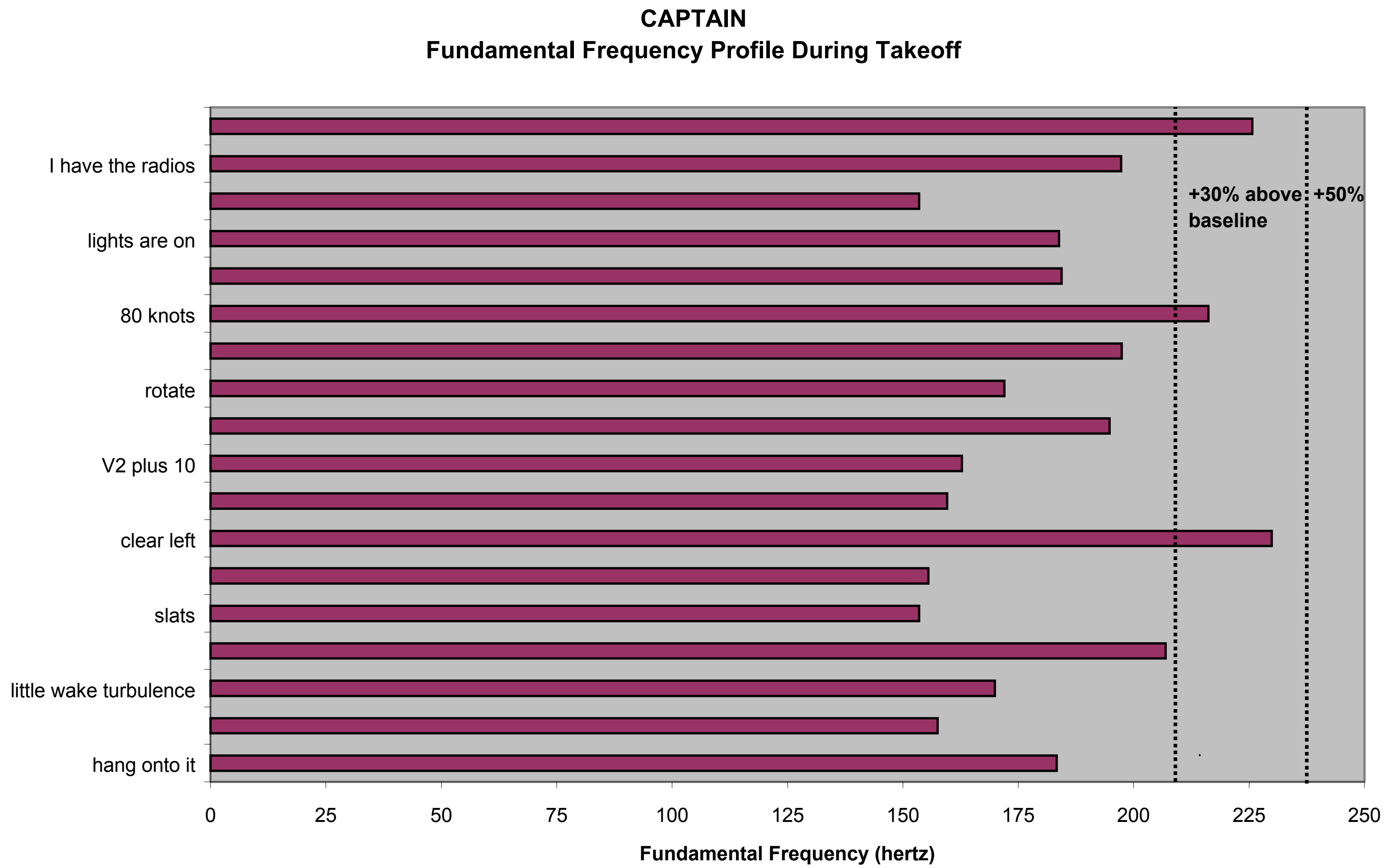
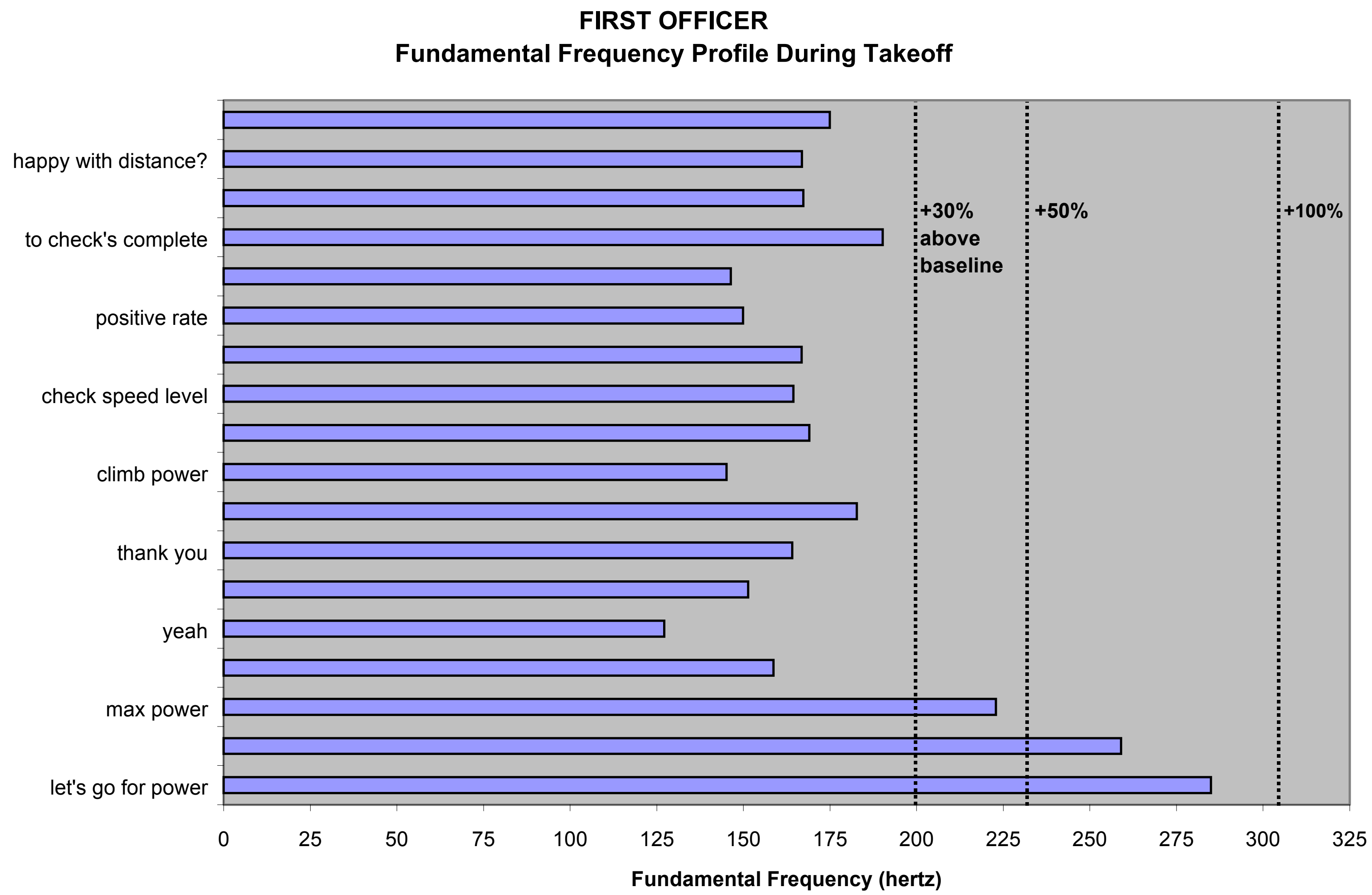

CAPTAIN
Fundamental Frequency Profile During Takeoff
I have the radios
lights are on
80 knots
rotate
V2 plus 10
clear left
slats
little wake turbulence
hang onto it
+30% above baseline
+50%
0
25
50
75
100
125
150
175
200
225
250
Fundamental Frequency (hertz)

FIGURE 3



FIRST OFFICER
Fundamental Frequency Profile During Takeoff
happy with distance?
to check's complete
positive rate
check speed level
climb power
thank you
yeah
max power
let's go for power
+30% above baseline
+50%
+100%
0
25
50
75
100
125
150
175
200
225
250
275
300
325
Fundamental Frequency (hertz)

FIGURE 4

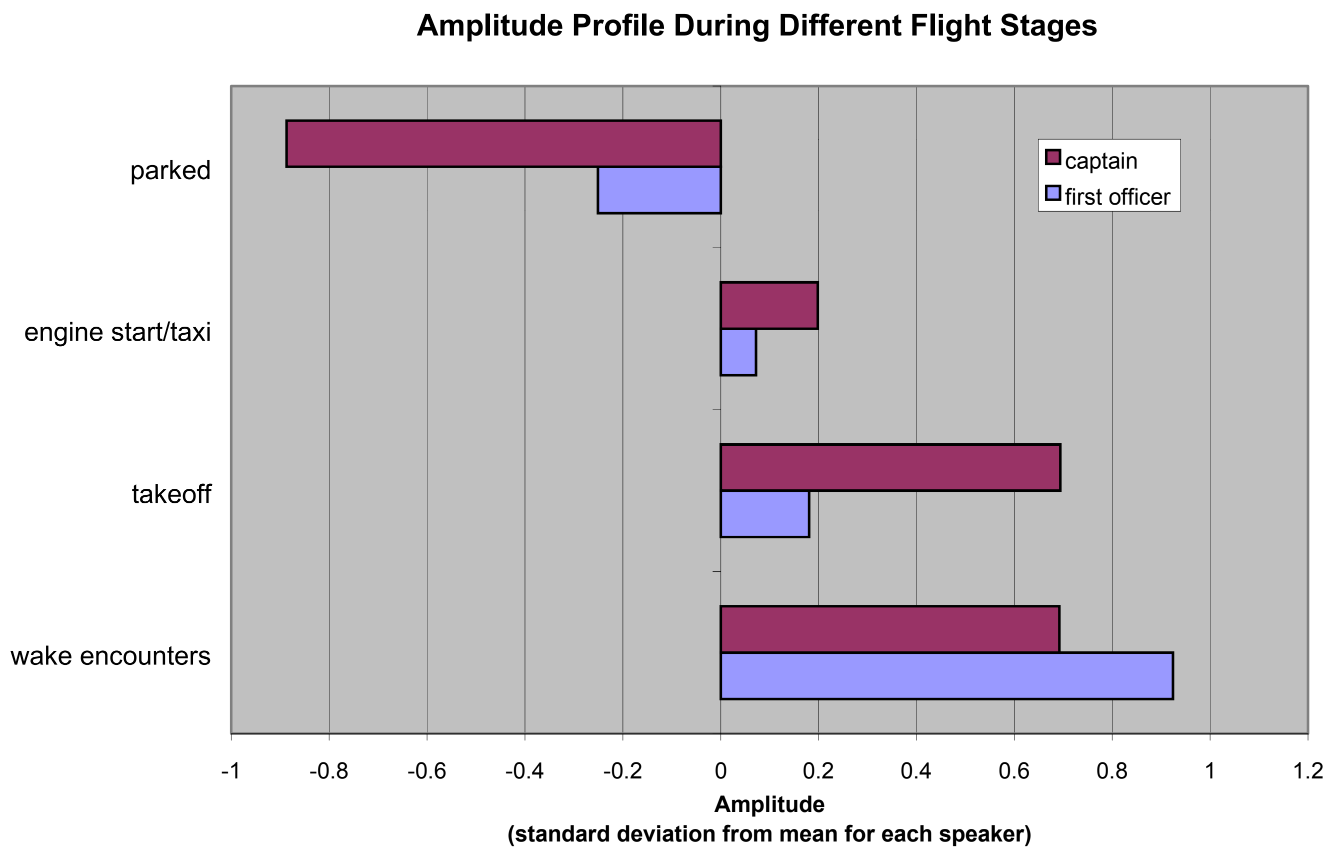

Amplitude Profile During Different Flight Stages
parked
engine start/taxi
takeoff
wake encounters
captain
first officer
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
1.2
Amplitude
(standard deviation from mean for each speaker)

FIGURE 5

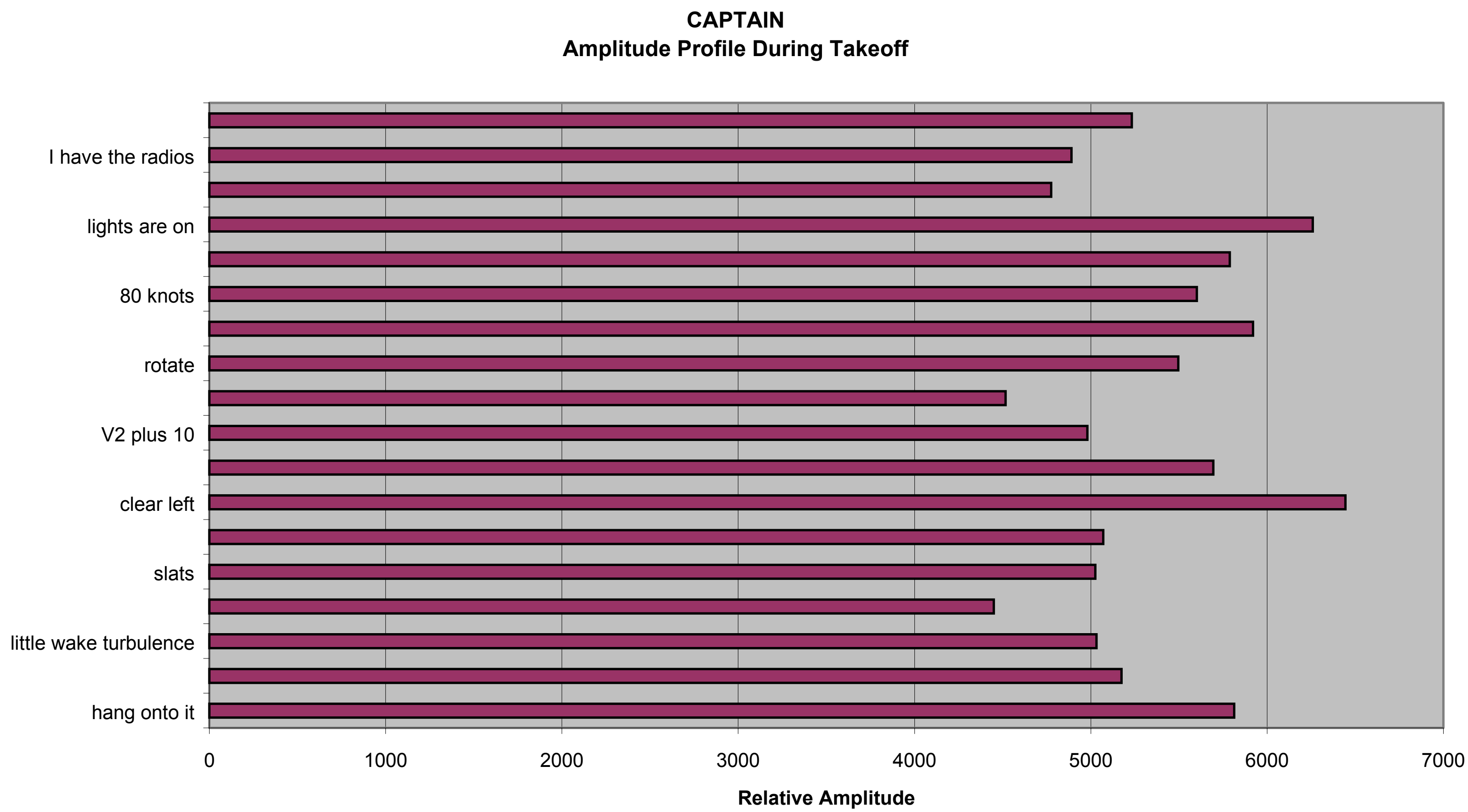
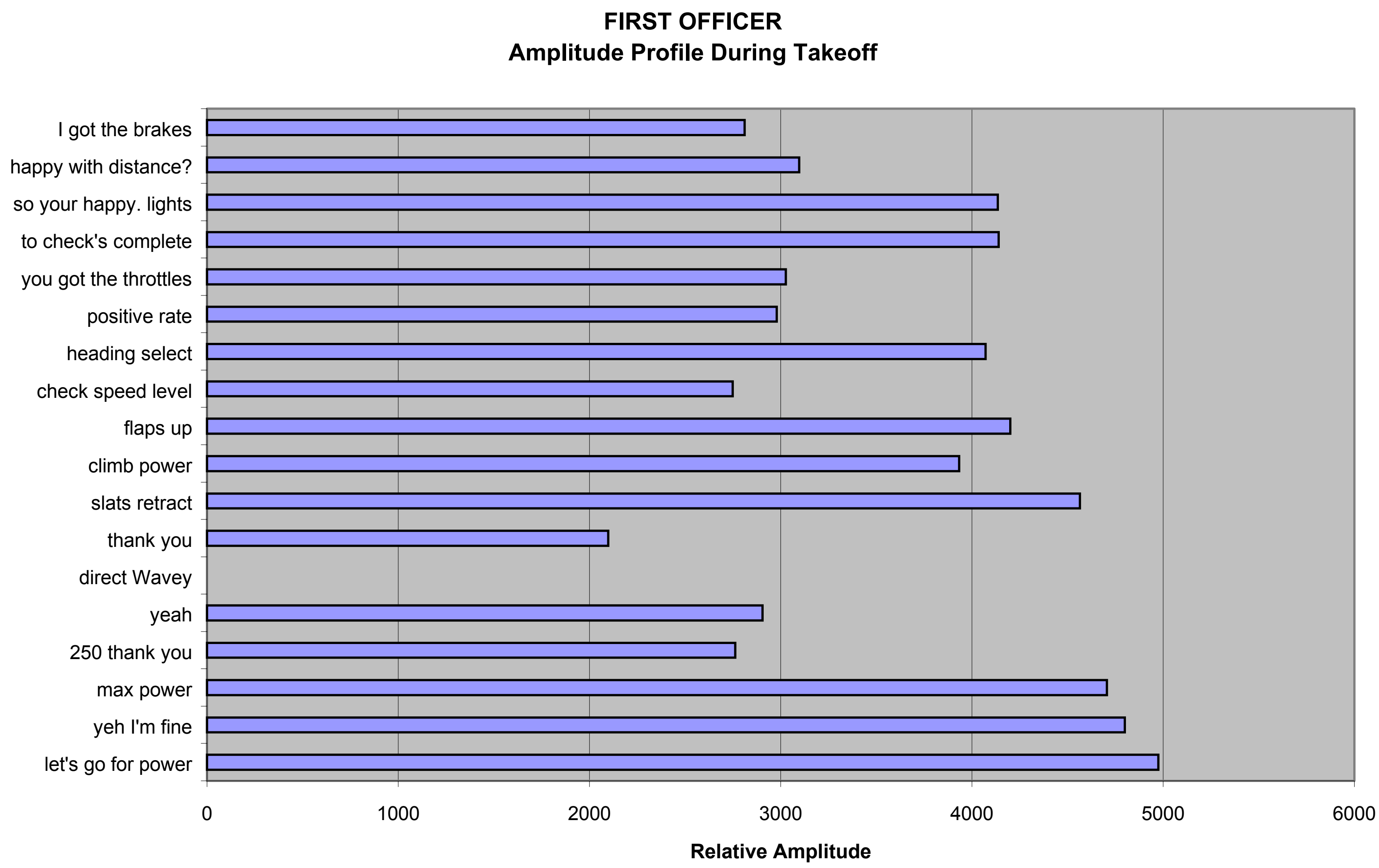

CAPTAIN
Amplitude Profile During Takeoff
I have the radios
lights are on
80 knots
rotate
V2 plus 10
clear left
slats
little wake turbulence
hang onto it
0
1000
2000
3000
4000
5000
6000
7000
Relative Amplitude

FIGURE 6


FIRST OFFICER
Amplitude Profile During Takeoff
I got the brakes
happy with distance?
so your happy. lights
to check's complete
you got the throttles
positive rate
heading select
check speed level
flaps up
climb power
slats retract
thank you
direct Wavey
yeah
250 thank you
max power
yeh I'm fine
let's go for power
0
1000
2000
3000
4000
5000
6000
Relative Amplitude